# A SEQUENTIAL PROCEDURE FOR SELECTING THE LARGEST OF k MEANS

by

H. Robbins, M. Sobel and N. Starr

University of California, University of Minnesota

and

Carnegie Institute of Technology

Joint Technical Report

April 10, 1967

Technical Report No. 88
Dept. of Statistics
University of Minnesota
Minneapolis, Minnesota

Technical Report No. 4
Dept. of Statistics
Carnegie Institute of Technology
Pittsburgh, Pennsylvania

# A SEQUENTIAL PROCEDURE FOR SELECTING THE LARGEST OF k MEANS

H. Robbins, M. Sobel, and N. Starr

University of California, University of Minnesota
and
Carnegie Institute of Technology

1. Introduction. Let $\pi_1, \pi_2, \ldots, \pi_k$ denote $k$ populations $(k \geq 2)$ in which we may observe the independent random variables $X_1, X_2, \ldots, X_k$, respectively, where $X_j$ is $N(\mu_j, \sigma^2)$ for $j = 1, 2, \ldots, k$. The $k + 1$ parameters $\mu_1, \mu_2, \ldots, \mu_k, \sigma^2$ are assumed unknown. We denote the ordered $\mu$-values by $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$, and if $\mu_{[k]} > \mu_{[k-1]}$ we refer to the population $\pi$ with $\mu = \mu_{[k]}$ as the best population. Our goal is to select the best population with probability at least $P^*$ whenever $\mu_{[k]} - \mu_{[k-1]} \geq \delta^*$; here $P^*$ and $\delta^*$ are preassigned constants with $1/k < P^* < 1$ and $\delta^* > 0$. In other words, letting CS denote correct selection and letting $\Omega_1$ be the set of all vectors $\vec{\omega} = (\mu_1, \mu_2, \ldots, \mu_k, \sigma^2)$ with $\mu_{[k]} - \mu_{[k-1]} \geq \delta^*$, we wish to obtain a procedure for which

$$(1) \qquad P(CS) \geq P^* \qquad \text{for all } \vec{\omega} \epsilon \Omega_1.$$

If $\sigma^2$ were known we could proceed as in [1]. Take a fixed number $n$ of independent observations on each of the $k$ random variables, and denote by $\bar{x}_j(n)$ the sample mean of the $n$ observations on $X_j (j = 1, 2, \ldots, k)$. If $\alpha$ is (say) the smallest $j$ such that $\bar{x}_j(n) \geq \bar{x}_i(n)$ $(i = 1, 2, \ldots, k)$, select $\pi_\alpha$.

Let $\Phi$ denote the c.d.f. of a $N(0, 1)$ random variable. From [1], for all $\vec{\omega} \epsilon \Omega_1$ and $n \geq 1$

$$(2) \qquad P_n(CS) \geq \int_{-\infty}^{\infty} \Phi^{k-1}\left(y + \frac{\delta^* \sqrt{n}}{\sigma}\right) d\Phi(y),$$

with equality when $\mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta^*$. Define for $\delta^* > 0$

(3) $$c \geq c(k, \delta^*) = (\frac{\delta^*}{h})^2$$

where $h = h(k, P^*)$ is the solution of

$$\int_{-\infty}^{\infty} \Phi^{k-1}(y + h)d\Phi(y) = P^*;$$

for $P^* > 1/k$ it is clear that $h > 0$ and hence $c > 0$. Then from (2) it is easily seen that (1) is satisfied provided $n$ is chosen so that

(4) $$n \geq \sigma^2/c.$$

Considering a continuous version of n, we denote by $n^* = \sigma^2/c$ the minimal fixed number of k-tuples of observations. Tables of $h$ for various values of $P^*$ and $k$ are given in [1].

Clearly, when $\sigma^2$ is unknown no fixed sample size procedure will work for all $\vec{\omega}\epsilon\Omega_1$. A two-stage procedure for this problem has been studied [2]. This procedure, while guaranteeing (1) for all $\vec{\omega}\epsilon\Omega_1$, is inefficient, since it utilizes only part of the sample to estimate $\sigma^2$. Accordingly, we consider here a more fully

<u>sequential procedure</u>: Let $x_{ij}$ denote the $i^{\underline{th}}$ observation on $X_j (j = 1, 2, \ldots, k)$, and define for $r \geq 2$

$$V_r = \frac{1}{k(r-1)} \sum_{j=1}^{k} \sum_{i=1}^{r} (x_{ij} - \bar{x}_j(r))^2.$$

Sampling proceeds sequentially where at each stage we make a single observation on each of the $k$ random variables $X_j$ and recompute a new estimate $U_r$ of $\sigma^2$. We terminate sampling at the $N^{\underline{th}}$ stage, where the random variable $N$ is the least <u>odd</u> integer $n \geq 5$ such that

(5) $$V_n \leq cn. \qquad \text{(cf.(4))}.$$

If $\alpha$ is (say) the smallest $j$ such that $\bar{x}_j(N) \geq \bar{x}_i(N)$ for all $i = 1, 2, \ldots, k$, we select $\pi_\alpha$. We require that sampling may be terminated only with an odd number of k-tuples of observations in order to simplify the later computations of the probability distribution of $N$.

2. Some Properties of the Procedure. As in [5, 6] we note that for each fixed $n$ the vector $\vec{X}(n) = (\bar{X}_1(n), \bar{X}_2(n), \ldots, \bar{X}_k(n))$ is independent of the event $\{N = n\}$. Accordingly, since the selection procedure depends only on $\vec{X}(N)$, we have from (2) for all fixed $\delta^* > 0$, $\vec{\omega} \epsilon \Omega_1$

$$
(6) \quad \begin{aligned} P(CS) &= \sum_n P(CS|N = n)\, P(N = n) = \sum_n P_n(CS)\, P(N = n) \\ &\geq \sum_n \int_{-\infty}^{\infty} \Phi^{k-1}\left(y + \frac{\delta^* \sqrt{n}}{\sigma}\right) d\Phi(y) \cdot P(N = n) = E\left\{\Phi^{k-1}\left(Y + \frac{\delta^* \sqrt{N}}{\sigma}\right)\right\} = \beta(\lambda), \text{say}, \end{aligned}
$$

where

$$
(7) \qquad \lambda = \sigma/\delta^*.
$$

That the procedure is asymptotically satisfactory as $\delta^* \to 0$ follows from [3, 4]. Let $\vec{\omega}$ be fixed and let $\delta^* \to 0$. Then $n^* \to \infty$, $N \to \infty$ a.s.,

$$
(8) \qquad N/n^* \to 1 \ \text{a.s.}, \ \frac{\delta^* \sqrt{N}}{\sigma} \to h \ \text{a.s.}, \ EN/n^* \to 1,
$$

and from (6), (8), and the Helly-Bray theorem

$$
(9) \qquad \liminf P(CS) \geq P^*, \ \vec{\omega} \epsilon \Omega_1.
$$

In fact, (8) and (9) hold without the assumption of normality provided only that the $X_j$ have finite fourth moments; even this condition can be relaxed.

As a measure of the cost of ignorance of $\sigma^2$ we define for fixed $\delta^* > 0$, $0 < \sigma^2 < \infty$ the quantity

$$
I = I(\lambda) = EN - n^*.
$$

This cost is negligible, as may be seen from the following

Theorem. For all $\delta^* > 0$, $0 < \sigma^2 < \infty$

$$
(10) \qquad I \leqq 5.
$$

Proof. Define

$$
U_r = \sum_{j=1}^{k} \sum_{i=1}^{r} (X_{ij} - \mu_j)^2;
$$

then for all $r \geq 1$,

(11) $$k(r-1)\, V_r \leq U_r \leq U_{r+2}.$$

Since from (5) for all odd $N > 5$

$$c(N-2) < V_{N-2}$$

we have from (11) for odd $N > 5$

$$c(N-2) < \frac{U_N}{k(N-3)} .$$

Thus for all odd $N \geq 5$

(12) $$ckN(N-5) < U_N.$$

Taking expectations on both sides of (12) we obtain by Wald's lemma

$$ckE\{N(N-5)\} < \sigma^2 k\ EN;$$

hence

$$(EN)^2 - 5\ EN < EN^2 - 5\ EN < \frac{\sigma^2}{c}\ EN,$$

and

$$EN < \frac{\sigma^2}{c} + 5 = n^* + 5.$$

In our proof we tacitly assumed that $EN < \infty$ so as to be able to apply Wald's lemma. This condition certainly holds if the $X_j$ are normal [see (3)]. However, we may eliminate the distributional assumption by the following device. Define for each $m > 5$ the random variable $N_m = \min(m, N)$ and deduce as before that $EN_m < n^* + 5$. Letting $m \to \infty$ we have $EN_m \uparrow EN \leq n^* + 5$. Thus (10) holds whatever the distribution of the $X_j$, provided only that $0 < \sigma^2 < \infty$.

We further observe that the inequality (10) cannot be sharpened since our procedure requires that $N \geq 5$, and accordingly (from (3)), $I \to 5$ as $\delta^* \to \infty$.

3. Small Sample Performance. We have established that the sequential procedure is asymptotically consistent and efficient (in the sense of [3]) and that the cost of ignorance of $\sigma^2$ is of little consequence when the sequential procedure is used, for all $\delta^* > 0$, $0 < \sigma^2 < \infty$. It remains to verify that P(CS) is

approximately $\geq P^*$, for all values of $\mu_1, \mu_2, \ldots, \mu_k$ for which $\vec{\omega}\varepsilon\Omega_1$ and for a wide range of values of the parameter $\lambda$ defined by (7). (c.f. Table I.)

Let $\{T_i\}$ $(i = 1, 2, \ldots)$ be independent random variables each with a chi-squared distribution with $k$ degrees of freedom. By Helmert's transformation we can write for all $n \geq 2$

$$\frac{k(n-1)V_n}{\sigma^2} = \sum_{i=1}^{n-1} T_i. \tag{13}$$

Let $n = 2m + 1$; then from (5) $N$ is the least integer $m \geq 2$ such that

$$V_{2m+1} \leq c(2m+1). \tag{14}$$

From (13) (recalling the definition (3) of c) we may rewrite (14) in the form

$$\sum_{n=1}^{2m} T_i \leq \frac{k(2m+1)2m}{h^2\lambda^2}.$$

The random variable $W = \frac{T_1 + T_2}{2}$ is the sum of $k$ independent standardized exponential random variables, so that $N$ is the least integer $m \geq 1$ such that

$$W_1 + W_2 + \ldots + W_m \leq a_{m+1};$$

here the $W_i$ are independently distributed as the sum of $k$ standardized exponential random variables and the constants $\{a_m\}$ are given by

$$a_2 = 0, \; a_m = \frac{k(2m-1)(m-1)}{h^2\lambda^2} \quad \text{for} \quad m = 3, 4, \ldots.$$

Thus for $0 < \lambda < \infty$ and $m = 1, 2, \ldots$ the distribution of N is given by

$$p_m(\lambda) = P_\lambda(N = 2m+1) = P_\lambda(Z_1 > a_2,\; Z_2 > a_3, \ldots,\; Z_{m-1} > a_m,\; Z_m \leq a_{m+1})$$

where for $m \geq 1$

$$Z_m = W_1 + W_2 + \ldots + W_m.$$

The following recursive scheme, which generalizes [5, 6], gives a method for computing $p_m(\lambda)$ $(m = 1, 2, \ldots)$ for given values of the parameter $\lambda$.

Define for $\alpha = 0, 1, \ldots, k-1$

$$h_k^{(\alpha)}(x) = \frac{x^{k-1-\alpha}}{(k-1-\alpha)!}$$

and for $m = 2, 3, 4, \ldots$

$$h_{mk}^{(\alpha)}(x) = \sum_{j=1}^{m-1} \sum_{\alpha=0}^{k-1} \frac{(x-a_m)^{jk+\alpha}}{(jk+\alpha)!} h_{(m-j)k}^{(\alpha)}(a_m).$$

Let $c_1 = c_2 = 1$, and for $m = 3, 4, \ldots,$ define

$$c_m = e^{-a_m} \sum_{j=0}^{m-1} \sum_{\alpha=0}^{k-1} h_{(m-j)k}^{(\alpha)}(a_m)$$

then for all $m \geq 1$

$$p_m = c_m - c_{m+1}.$$

Thus from (6), for any fixed value of the parameter $\lambda$

$$P(CS;\lambda) \geq p_m(\lambda) \int_{-\infty}^{\infty} \Phi(y + \frac{\sqrt{2m+1}}{\lambda}) d\Phi(y) = \beta(\lambda),$$

for all $\mu_1, \mu_2, \ldots, \mu_k$ for which $\vec{\omega} \epsilon \Omega_1$. Moreover, the expected sample size $EN = E_\lambda N$ is defined by

$$EN = \sum_{m=1}^{\infty} (2m+1)\, p_m(\lambda) = 2 \sum_{m=1}^{\infty} mp_m(\lambda) + 1.$$

Exact computations of $\beta$ and $EN$ have been carried out (for a number of values of $\lambda$) when $P^* = .95$, $k = 2, 3, 4, 5$, and are presented in the accompanying Table I. The values of the fixed sample size $n^* = \sigma^2/c = h^2\lambda^2$ which would be used if $\sigma^2$ were known are included for comparison with $EN$.

We have no proof that the minimum value of the lower bound $\beta$ of $P(CS)$ for $\vec{\omega} \epsilon \Omega_1$ is attained in the computed range $n^* = 2, \ldots, 80$, but this appears to be the case. If this is so, then Table I shows that $\beta$ never gets much below $P^* = .95$.

Remark: The method used in proving (10) can also be applied to the problem of obtaining an upper bound on the expected sample size for the problem considered in [3, 4, 6].

4. Acknowledgement. The authors wish to thank Professor M. M. Rao of Carnegie Institute of Technology for a helpful discussion, Mr. Y. L. Tong of the University of Minnesota for reading the manuscript and Mrs. Elaine Frankowski of the University of Minnesota for helping with the computations in Table I.

TABLE I

Exact Computations of $\beta$ and EN for $P^* = .95$

(as a function of $n^*$ and k)

| $n^*=h^2\lambda^2$ | k = 2 | | k = 3 | | k = 4 | | k = 5 | |
|---|---|---|---|---|---|---|---|---|
| | EN | $\beta$ | EN | $\beta$ | EN | $\beta$ | EN | $\beta$ |
| 2 | 5.02 | .99539 | 5.01 | .99763 | 5.00 | .99840 | 5.10 | .99879 |
| 4 | 5.63 | .97223 | 5.45 | .97509 | 5.46 | .97636 | 5.42 | .97710 |
| 6 | 6.88 | .95503 | 6.87 | .95598 | 6.86 | .95682 | 6.86 | .95759 |
| 8 | 8.44 | .94620 | 8.56 | .94734 | 8.64 | .94918 | 8.70 | .95099 |
| 10 | 10.17 | .94217 | 10.42 | .94431 | 10.57 | .94713 | 10.67 | .94956 |
| 12 | 12.01 | .94065 | 12.37 | .94376 | 12.56 | .94708 | 12.67 | .94964 |
| 14 | 13.92 | .94043 | 14.25 | .94420 | 14.57 | .94760 | 14.69 | .95000 |
| 16 | 15.87 | .94084 | 16.36 | .94497 | 16.59 | .94821 | 16.70 | .95033 |
| 18 | 17.86 | .94155 | 18.38 | .94576 | 18.60 | .94872 | 18.71 | .95055 |
| 20 | 19.87 | .94236 | 20.40 | .94648 | 20.62 | .94911 | 20.73 | .95068 |
| 25 | 25.00 | .94428 | 25.46 | .94779 | 25.65 | .94969 | 25.74 | .95079 |
| 30 | 30.00 | .94576 | 30.49 | .94853 | 30.67 | .94993 | 30.75 | .95077 |
| 35 | 35.07 | .94680 | 35.52 | .94896 | 35.68 | .95003 | 35.76 | .95072 |
| 40 | 40.12 | .94751 | 40.53 | .94921 | 40.68 | .95009 | 40.76 | .95066 |
| 45 | 45.16 | .94801 | 45.55 | .94937 | 45.69 | .95011 | 45.76 | .95061 |
| 50 | 50.20 | .94835 | 50.55 | .94947 | 50.69 | .95012 | 50.76 | .95057 |
| 60 | 60.24 | .94879 | 60.56 | .94961 | 60.70 | .95012 | 60.77 | .95049 |
| 70 | 70.26 | .94904 | 70.57 | .94969 | 70.70 | .95012 | 70.77 | .95043 |
| 80 | 80.28 | .94921 | 80.57 | .94974 | 80.70 | .95011 | 80.77 | .95039 |

## BIBLIOGRAPHY

[1] Bechhofer, R.E. (1954). "A single-sample multiple decision procedure for ranking means of normal populations with known variances," Ann. Math. Statist., 25, 16-39.

[2] Bechhofer, R.E., Dunnett, C.W., and Sobel, M. (1954). "A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance," Biometrika 41, 170-176.

[3] Chow, Y.S. and Robbins, H. (1965). "On the asymptotic theory of fixed-width sequential confidence intervals for the mean," Ann. Math. Statist., 36, 463-467.

[4] Gleser, L. J., Robbins, H., and Starr, N. (1964). "Some asymptotic properties of fixed-width sequential confidence intervals for the mean of a normal population with unknown variance," Columbia Univ. Report.

[5] Robbins, H. (1959). "Sequential estimation of the mean of a normal population," Probability and Statistics--The Harald Cramér Volume, 235-245, Almqvist and Wiksell, Uppsala, Sweden.

[6] Starr, N. (1966). "The performance of a sequential procedure for the fixed-width interval estimation of the mean," Ann. Math. Statist., 37, 36-50.